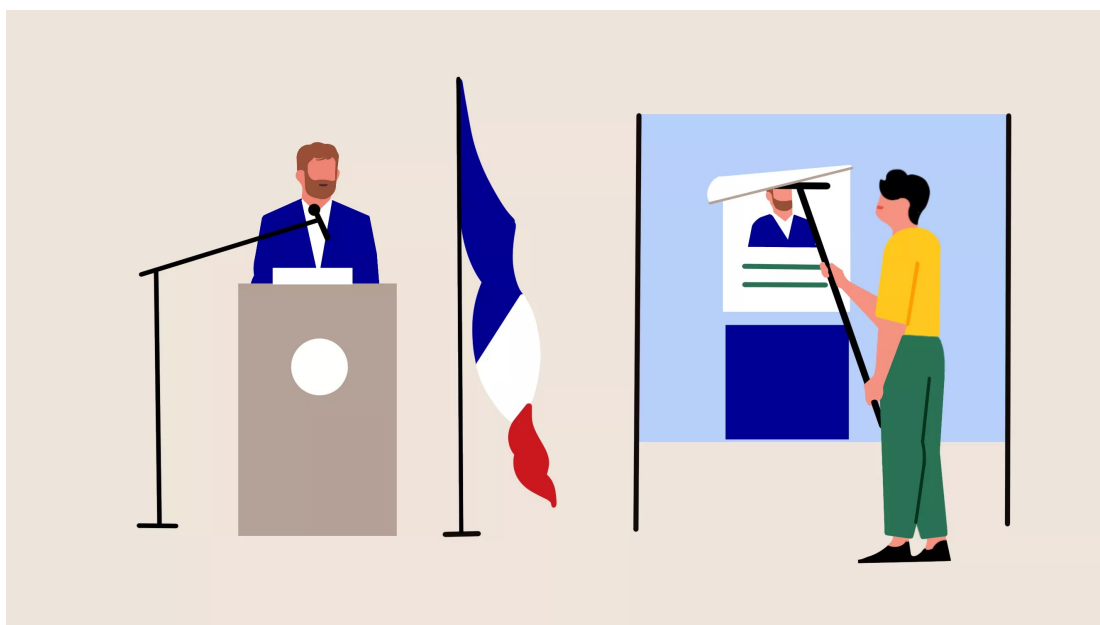


RAPPORT DE PROJET BIG DATA

GÉOGRAPHIE ET DÉTERMINANTS DU VOTE PRÉSIDENTIEL FRANÇAIS 2022
Analyse multi-échelle et socio-économique via Big Data

Par : Mathis Telle



Sous la direction de Mr. Devan Sohier
Remis le : 15 mai 2025

Plan du rapport

Introduction et cadre de l'étude

Contexte, objectifs et jeux de données (Ministère de l'Intérieur & INSEE)
Méthodes mobilisées et organisation du rapport

I - Analyse nationale

Scores des candidats
Taux de participation, abstention, votes blancs/nuls
Principales variations régionales

II - Analyse départementale (avec focus Outre-mer)

Classement des départements par score et écart entre candidats
Zoom sur la participation et le candidat dominant en Outre-mer

III- Analyse circonscriptionnelle et facteurs socio-économiques

Enrichissement INSEE
ACP pour décrire les grands axes socio-économiques
Régressions linéaires :
modèles « socio-éco seuls »
modèles enrichis
modèles « tranches d'âge »

IV - Conclusion et perspectives

Introduction et cadre de l'étude

L'élection présidentielle française constitue un moment majeur de la vie démocratique, révélant non seulement les préférences des citoyens, mais aussi les dynamiques sociales, économiques et territoriales. À l'ère du Big Data, la disponibilité d'informations des bureaux de vote jusqu'aux circonscriptions législatives offre une opportunité unique d'analyser les résultats électoraux à la fois sous l'angle géographique et socio-économique. Ce rapport, réalisé dans le cadre du projet de Big Data, mobilise les données officielles du premier tour de la présidentielle 2022 (Ministère de l'Intérieur) et les indicateurs socio-économiques par circonscription (INSEE), afin de mettre en œuvre les méthodes vues en cours : pivotements de données, analyses en composantes principales, régressions, etc.. et d'en tirer des interprétations sur ce vote.

QUESTIONS DE RECHERCHE:

Ce travail s'organise autour de trois questions centrales :

- Quels sont les facteurs socio-économiques et démographiques qui structurent la répartition des suffrages ?
- Comment ces facteurs se manifestent-ils à différentes échelles (nationale, régionale, départementale, circonscriptionnelle) et quelles disparités émergent entre Métropole et Outre-mer ?
- Quelle valeur ajoutée offrent les méthodes multivariées (ACP, clustering, régressions linéaires) pour quantifier et interpréter l'impact respectif de ces variables sur les choix électoraux ?

L'étude se déploie en trois grandes parties, suivant une démarche que l'on pourrait qualifier de décroissante.

Tout d'abord une analyse nationale, où l'on dresse un panorama global des scores des candidats ainsi que les taux de participation, d'abstention, etc.. à l'échelle nationale (métropolitaine et ultramarine).

Puis une analyse départementale, qui affine l'examen à l'échelle des départements français, identifie les territoires de prédilection de chaque candidat et propose un focus spécifique sur les Outre-mer, où les enjeux socio-économiques et le comportement électoral présentent des caractéristiques singulières.

Et enfin une analyse circonscriptionnelle et socio-économique, qui intègre les données détaillées de l'INSEE pour interroger les déterminants structurels du vote : pauvreté, chômage, niveau de diplôme... À l'aide d'analyses multivariées (ACP, régressions, ou encore clustering), nous cherchons à éclairer les corrélations entre contextes territoriaux et scores électoraux, tout en restant vigilants aux limites d'interprétation. L'objectif est de comprendre dans quelle mesure ces facteurs structurent les comportements de vote en comparant les enseignements de l'ACP et des modèles de régression. Ce rapport n'a pas pour objet d'expliquer exhaustivement les comportements électoraux individuels ce serait l'apanage des travaux de sciences sociales de longue haleine mais d'essayer de montrer la compréhension des méthodes analytiques vues en cours couplée à celle que j'ai pu acquérir en dehors afin de produire des insights pertinents sur un phénomène politique d'ampleur nationale. Les résultats, présentés pour chaque niveau d'analyse, seront commentés et mis en perspective pour en dégager les principaux enseignements.

I- Analyse nationale

I.1 Données et prétraitements

Les résultats du premier tour de la présidentielle 2022, provenant du ministère de l'Intérieur, ont été importés puis nettoyés afin de garantir la fiabilité des analyses. [1]

Celui-ci comporte, pour chaque région :

17 colonnes « globales » (inscrits, votants, abstentions, exprimés, etc.)

Après centrage-réduction des formats et vérification des types, nous avons transformé le schéma initial « blocs candidats » en un format « une ligne par région x candidat », permettant d'isoler les pourcentages de voix exprimées pour Emmanuel Macron, Marine Le Pen et Jean-Luc Mélenchon.

I.2 Participation électorale : abstention, blancs et nuls

L'objectif de cette sous-partie est de quantifier la mobilisation afin d'évaluer le poids des électeurs non exprimés pour mieux comprendre la légitimité du scrutin.

Nous avons donc procédé comme ceci:

- Extraction des taux : à partir des 17 premières colonnes, nous avons isolé le taux d'abstention (% Abs/Ins), le taux de participation (% Vot/Ins), les votes blancs et nuls (% Blancs/Ins, % Nuls/Ins).

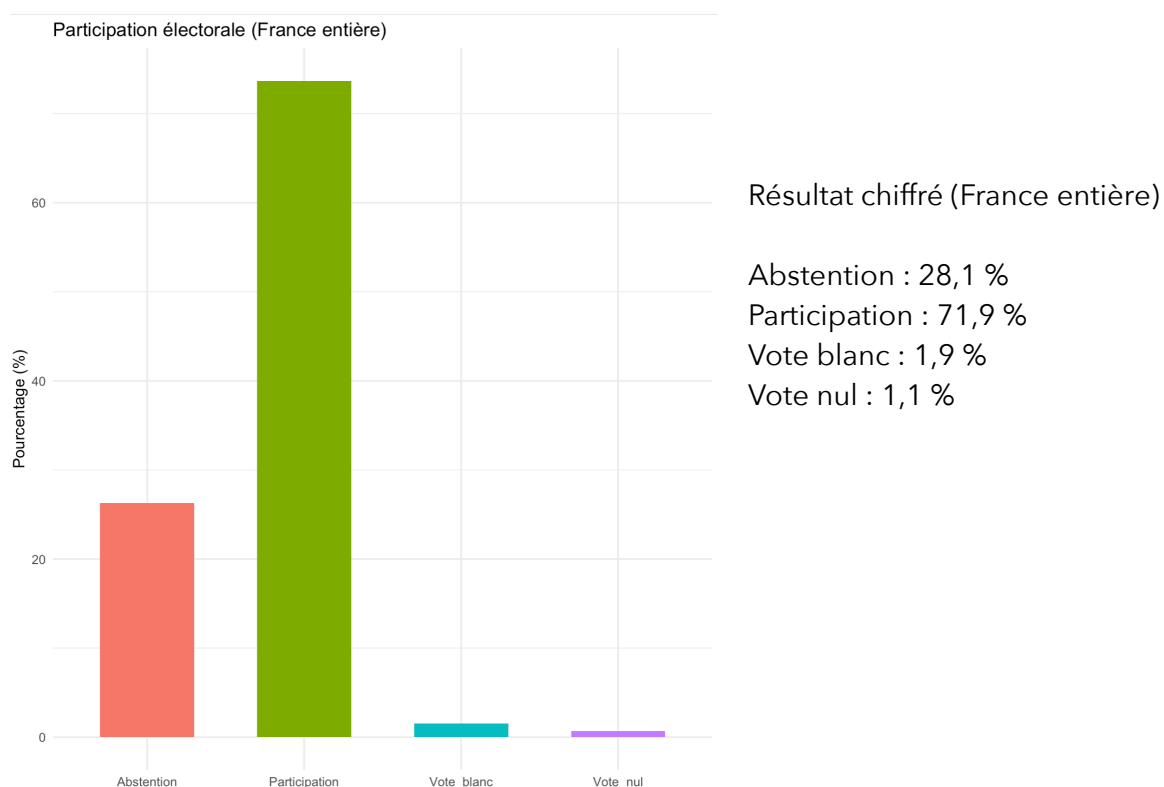


Figure 1 - Participation électorale France entière

Taux de participation	1 ^{er} tour		
	En 2012	En 2017	Différence
à midi	28,29 % ²²³	28,54 % ²²⁴	▲ 0,25
à 17 heures	70,59 % ²²⁷	69,42 % ²²⁴	▼ 1,17
final	79,48 % ²²⁹	77,77 % ²³⁰	▼ 1,71

Figure 2 - Participation électorale France entière aux deux élections précédentes.

Le taux de participation atteint 71,9 % des électeurs inscrits, traduisant une abstention de 28,1 %, soit une hausse de 6,1 points par rapport à 2017 (22,0 %) et de 8,5 points depuis 2012 (19,6 %). Les votes blancs et nuls, respectivement à 1,9 % et 1,1 %, atteignent également des niveaux inédits depuis dix ans, illustrant une défiance croissante envers l'offre politique. (Wikipédia) Cependant, ces chiffres, agrégés au niveau régional, masquent des disparités locales (et ne rendent pas compte des dynamiques intra-régionales. Cette forte volatilité de la participation et de la protestation fonde le contexte pour l'analyse des scores des candidats : nous examinerons désormais comment ces derniers se répartissent les suffrages exprimés.

1.3 Scores des candidats : chiffres bruts et pourcentages

Objectif : identifier les principaux compétiteurs

nous avons d'abord calculé, pour chacun des 12 candidats, le total de voix nationales et le pourcentage des voix exprimées, puis classé ces candidats par ordre décroissant. Cette opération a permis d'identifier de façon objective les trois mieux placés :

	N_Panneau	Sexe	Nom	Prenom	Voix	Pourc_Voix_Ins	Pourc_Voix_Exp
	<dbl>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
1	3	M	MACRON	Emmanuel	9783058	20.1	27.8
2	5	F	LE PEN	Marine	8133828	16.7	23.2
3	7	M	MÉLENCHON	Jean-Luc	7712520	15.8	22.0
4	6	M	ZEMMOUR	Éric	2485226	5.1	7.07
5	10	F	PÉCRESSE	Valérie	1679001	3.44	4.78
6	9	M	JADOT	Yannick	1627853	3.34	4.63
7	4	M	LASSALLE	Jean	1101387	2.26	3.13
8	2	M	ROUSSEL	Fabien	802422	1.65	2.28
9	12	M	DUPONT-AIGNAN	Nicolas	725176	1.49	2.06
10	8	F	HIDALGO	Anne	616478	1.26	1.75
11	11	M	POUTOU	Philippe	268904	0.55	0.77
12	1	F	ARTHAUD	Nathalie	197094	0.4	0.56

Figure 3 - Classement national des candidats par nombre de voix

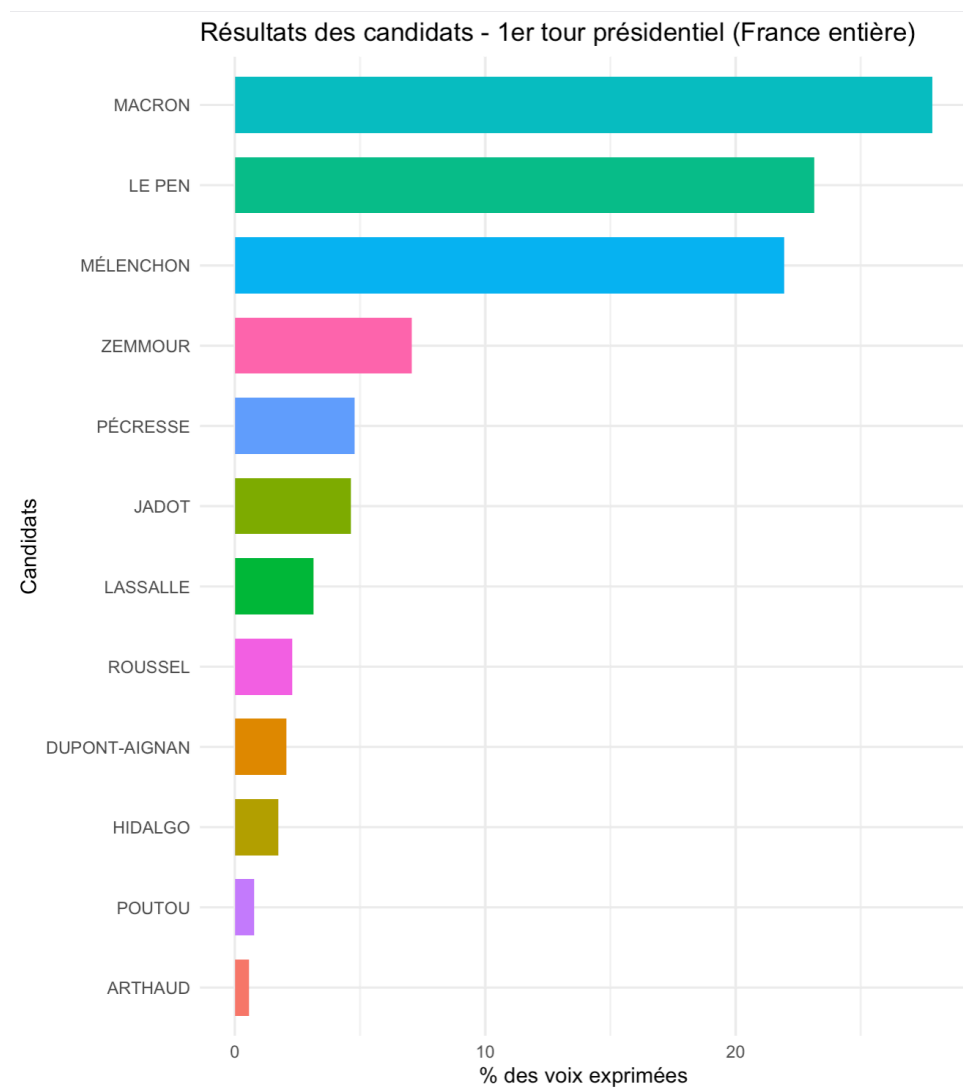


Figure 4 - Classement national des candidats par % de voix exprimées.

Cette démarche a permis de faire émerger Emmanuel Macron (9,78 millions de voix), Marine Le Pen (8,13 millions), et Jean-Luc Mélenchon (7,72 millions), comme les trois acteurs majeurs du premier tour. En concentrant notre attention sur ces trois candidats, nous capturons 73 % des suffrages exprimés.

Cette sélection et ces scores vont alors constituer le socle sur lequel s'appuieront toutes les analyses spatiale et socio-économique que nous ferons par la suite. D'une part, l'analyse spatiale comparera les territoires (régions et départements) où chacun de ces trois leaders a creusé son avantage, afin de dégager les grands clivages géographiques du scrutin.

Ces résultats nationaux dressent un panorama général, mais masquent d'importantes disparités locales. La Partie II propose un décryptage plus fin, d'abord à l'échelle des régions, puis des départements (avec un focus sur les Outre-mer), afin de mettre en évidence les territoires où chaque leader concentre ses soutiens.

II - Analyse départementale : régions, départements et Outre-mer

II.1 Analyse au niveau des régions/départements

Pour explorer la répartition géographique des trois principaux candidats, nous avons d'abord récupéré les résultats par région dans le fichier Excel [3]. Comme précédemment, les 17 premières colonnes fournissent les totaux structurels, puis apparaissent pour chaque candidat sept variables (numéro de panneau, sexe, prénom, nom, voix, % voix/inscrits, % voix/exprimés). En s'appuyant sur un `pivot_longer()` suivi d'un `pivot_wider()`, nous avons extrait uniquement les pourcentages de voix exprimées obtenus par Emmanuel Macron, Marine Le Pen et Jean-Luc Mélenchon, et reformé un tableau à une ligne par région.

```
> top_lepen
# A tibble: 10 x 4
  `Libellé du département` Prénom Nom    `% Voix/Exp`
  <chr>                   <chr> <chr>      <dbl>
1 Mayotte                 Marine LE PEN    42.7
2 Aisne                   Marine LE PEN    39.2
3 Pas-de-Calais           Marine LE PEN    38.7
4 Haute-Marne             Marine LE PEN    36.6
5 Ardennes                Marine LE PEN    36.0
6 Meuse                   Marine LE PEN    35.1
7 Haute-Saône             Marine LE PEN    34.6
8 Aube                    Marine LE PEN    33.0
9 Somme                   Marine LE PEN    32.8
10 Pyrénées-Orientales    Marine LE PEN    32.7

> top_macron
# A tibble: 10 x 4
  `Libellé du département` Prénom Nom    `% Voix/Exp`
  <chr>                   <chr> <chr>      <dbl>
1 Français établis hors de France Emmanuel MACRON    45.1
2 Nouvelle-Calédonie      Emmanuel MACRON    40.5
3 Polynésie française     Emmanuel MACRON    40.2
4 Wallis et Futuna        Emmanuel MACRON    39.5
5 Hauts-de-Seine          Emmanuel MACRON    37.1
6 Mayenne                 Emmanuel MACRON    36.4
7 Vendée                  Emmanuel MACRON    35.6
8 Maine-et-Loire          Emmanuel MACRON    35.6
9 Paris                   Emmanuel MACRON    35.3
10 Ille-et-Vilaine        Emmanuel MACRON    34.5

> top_melenchon
# A tibble: 10 x 4
  `Libellé du département` Prénom Nom    `% Voix/Exp`
  <chr>                   <chr> <chr>      <dbl>
1 Guadeloupe              Jean-Luc MÉLENCHON    56.2
2 Martinique              Jean-Luc MÉLENCHON    53.1
3 Guyane                  Jean-Luc MÉLENCHON    50.6
4 Seine-Saint-Denis       Jean-Luc MÉLENCHON    49.1
5 Saint-Pierre-et-Miquelon Jean-Luc MÉLENCHON    40.9
6 La Réunion              Jean-Luc MÉLENCHON    40.3
7 Val-d'Oise              Jean-Luc MÉLENCHON    33.2
8 Val-de-Marne            Jean-Luc MÉLENCHON    32.7
9 Paris                   Jean-Luc MÉLENCHON    30.1
10 Saint-Martin/Saint-Barthélemy Jean-Luc MÉLENCHON    28.1
```

Figure 4 - Classements des régions dominées par les candidats principaux

Le classement des régions selon ces trois scores révèle immédiatement des logiques très distinctes :

Macron atteint son maximum dans les Français établis hors de France (45,1 %), suivi par la Nouvelle-Calédonie (40,5 %) et la Polynésie française (40,2 %).

Le Pen domine surtout dans Mayotte (42,7 %), l'Aisne (39,2 %), le Pas-de-Calais (38,7 %) et, de manière générale, dans plusieurs départements ruraux du nord-est.

Mélenchon est en tête en Guadeloupe (56,2 %), Martinique (53,1 %) et Guyane (50,6 %), mais marque également des scores remarquables en Seine-Saint-Denis (49,1 %) et Saint-Pierre-et-Miquelon (40,9 %).

Pour mieux comprendre ce que cela implique et évaluer les forces et faiblesse de chacun nous avons comparé les départements où chaque candidats dominait et comment ses adversaires s'en sortaient.

```
# Groups: Candidat_dominant [3]
```

`Libellé du département`	MACRON	`LE PEN`	MÉLENCHON	Candidat_dominant	Score_dominant
<chr>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>
1 Mayotte	16.9	42.7	24.0	LE PEN	42.7
2 Aisne	22.1	39.2	15.5	LE PEN	39.2
3 Pas-de-Calais	24.6	38.7	15.8	LE PEN	38.7
4 Haute-Marne	23.3	36.6	14.1	LE PEN	36.6
5 Ardennes	23.6	36.0	16.6	LE PEN	36.0
6 Français établis hors de France	45.1	5.29	21.9	MACRON	45.1
7 Nouvelle-Calédonie	40.5	18.8	13.8	MACRON	40.5
8 Polynésie française	40.2	19.5	13.3	MACRON	40.2
9 Wallis et Futuna	39.5	10.8	9.35	MACRON	39.5
10 Hauts-de-Seine	37.1	8.37	25.8	MACRON	37.1
11 Guadeloupe	13.4	17.9	56.2	MÉLENCHON	56.2
12 Martinique	16.3	13.4	53.1	MÉLENCHON	53.1
13 Guyane	14.2	17.7	50.6	MÉLENCHON	50.6
14 Seine-Saint-Denis	20.3	11.9	49.1	MÉLENCHON	49.1
15 Saint-Pierre-et-Miquelon	19.8	17.0	40.9	MÉLENCHON	40.9

```
>
```

Que peut on en déduire?

- Fortes concentrations géographiques

Le Pen domine dans plusieurs départements de la « diagonale du vide » ou la « diagonale des faibles densités » (*large bande du territoire français allant de la Meuse aux Landes où les densités de population sont relativement faibles par rapport au reste de la France*) et de la France rurale du nord-est (Mayotte fait ici exception car il s'agit du département d'outre-mer à très forte abstention et où son score est mécaniquement surélevé), avec des parts de voix comprises entre 36 % et 43 %. Macron l'emporte massivement dans trois collectivités « expatriées » (Français établis hors de France, Nouvelle-Calédonie, Polynésie) ainsi que dans des territoires à profil socio-économique élevé (Hauts-de-Seine), où il capte 37 % à 45 % des voix exprimées. Mélenchon lui, arrive en tête dans plusieurs DOM (Guadeloupe, Martinique, Guyane) et dans un département urbain défavorisé (Seine-Saint-Denis), avec des scores supérieurs à 49 %.

- Profil socio-politique des électorats

Les territoires d'outre-mer incluent à la fois des logiques d'abstention (p.ex. Mayotte) et une sensibilité très marquée aux thématiques sociales et économiques portées par Mélenchon (Martinique, Guadeloupe). Le basculement de Macron dans les territoires à forte proportion d'expatriés et dans les zones aisées souligne son ancrage dans un électorat plus mobile et plus diplômé. Le succès de Le Pen dans les départements ruraux et périurbains témoigne de l'écho de son discours souverainiste et anti-métropoles auprès d'électeurs marqués par le sentiment de déclassement

En synthèse, cette figure illustre que le vote des trois principaux candidats n'est pas homogène : il révèle des clivages territoriaux forts

II.2 Zoom « Outre-mer »

LES DÉPARTEMENTS ET COLLECTIVITÉS D'OUTRE-MER CONSTITUENT UN CAS PARTICULIER :

Abstention élevée, résultats incohérents parfois avec l'actualité, les tendances et les besoins des populations vivants dans ces départements. Comment expliquer cela Macron domine en Polynésie (40,2 %) et Nouvelle-Calédonie (40,5 %), Mélenchon écrase ses rivaux en Guadeloupe (56,2 %), Martinique (53,1 %) et Guyane (50,6 %), tandis que Le Pen reste majoritaire à Mayotte (42,7 %). Pourquoi autant d'abstention?

	`Libellé du département`	`% Abs/Ins`	Candidat_dominant	Score_dominant
	<chr>	<dbl>	<chr>	<dbl>
1	Polynésie française	69.1	MACRON	40.2
2	Nouvelle-Calédonie	66.6	MACRON	40.5
3	Guyane	63.8	MÉLENCHON	50.6
4	Mayotte	59.7	LE PEN	42.7
5	Martinique	57.3	MÉLENCHON	53.1
6	Guadeloupe	55.2	MÉLENCHON	56.2
7	La Réunion	46.4	MÉLENCHON	40.3
8	Saint-Pierre-et-Miquelon	45.0	MÉLENCHON	40.9
9	Wallis et Futuna	43.3	MACRON	39.5

Figure 6 - Score des candidats (Outre-mer)

zone	pauvreté	chômage	peu_diplomes	niveau_vie_median	pauvres_menages	aises
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 Hexagone	14.6	6	15.0	22259.	14.6	9.96
2 Outre-mer	33.7	11.5	31.6	16911.	33.7	7.89

Figure 7- indicateurs socio-économique (Outre-mer)

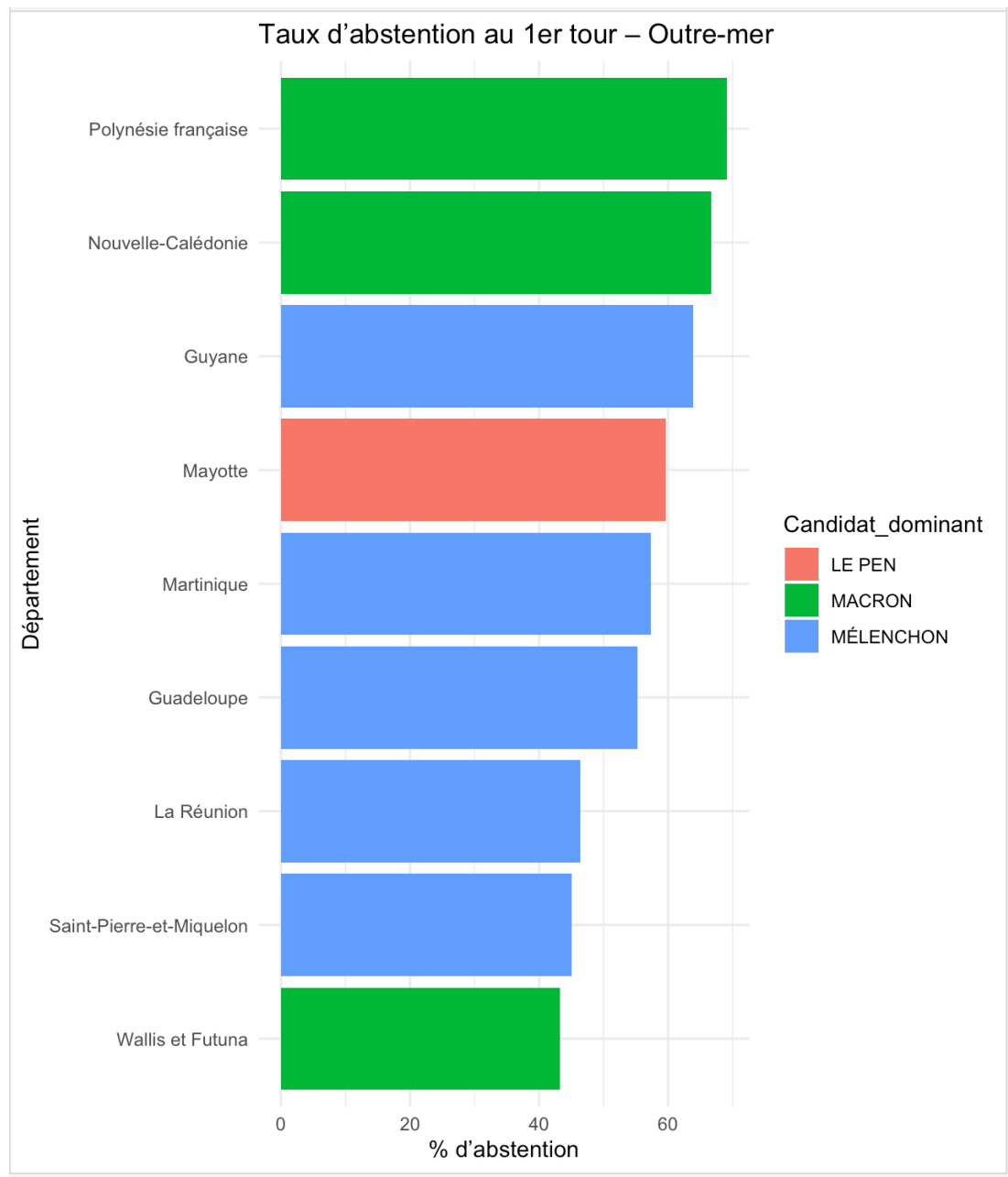


Figure 8 - Taux d'abstention (Outre-mer)

La géographie électorale s'appuie sur des travaux fondateurs qui lient clairement les contextes socio-économiques aux comportements de vote. Lipset & Rokkan (1967) - *Structures de clivages, systèmes de partis et alignement des électeurs* ont montré, dès les années 1960, que les clivages sociaux (statut socio-professionnel, niveau de diplôme) structurent durablement les préférences partisans. Plus récemment, Blais & Nadeau (1996) - *Measuring strategic voting: A two-step procedure* ont mis en évidence l'impact du chômage et de la pauvreté sur le taux de participation, tandis que Desage & Guérin-Pace (2005) ont cartographié les effets du vieillissement démographique sur les scores des partis en France. Enfin, Jackman & Miller (1995) soulignent l'importance du revenu médian et des conditions de logement comme déterminants de l'abstention et du vote protestataire.

Ces travaux justifient notre choix d'indices socio-économiques (chômage, pauvreté, diplôme) ainsi que de variables démographiques (âge moyen, quantiles) et logement (vacance, ancienneté du parc) : chacun est documenté dans la littérature comme un facteur explicatif majeur des dynamiques électorales.

Effectivement, les inégalités socio-économiques sont le moteur de l'abstention et des votes « contestataires ». Il y a un sentiment de déclassement et marginalisation dans ces départements et cela depuis des décennies. Avec un tiers des ménages ultramarins sous le seuil de pauvreté et un taux de chômage deux fois plus élevé qu'en métropole, beaucoup se sentent peu concernés par une élection présidentielle où les programmes nationaux paraissent déconnectés de leur quotidien. Les électeurs les plus précaires, souvent sur des emplois informels ou saisonniers, renoncent à « perdre » une demi-journée (voire plus) pour voter.

Pourquoi le candidat « Mélenchon » l'emporte là où la pauvreté et l'exclusion sont fortes?

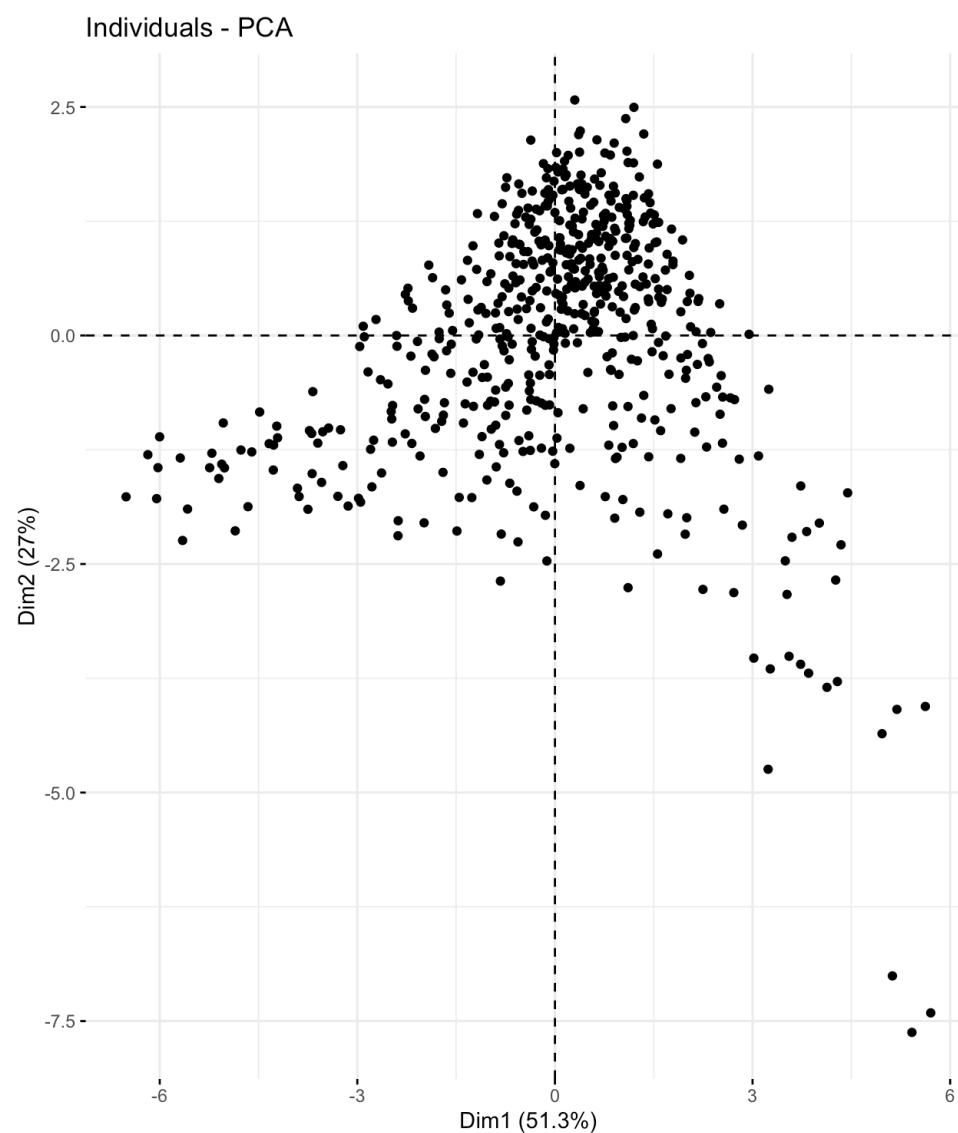
Le grand gagnant du suffrage en outremer reste le candidat Mélenchon qui sort en tête dans 5 des 9 départements. Et cela s'explique avec cette comparaison socio-économique. L'extrême précarité en Outre-mer apparaît clairement corrélée à la très forte poussée de ce candidat dans les « îles » où le ressentiment socio-économique est le plus marqué. Son programme bataille sur le blocage des prix, la hausse du RSA et la création d'emplois publics répond directement aux urgences économiques et sociales d'une population fragilisée.

En résumé, cette partie II établit d'abord les cartes de force des trois leaders à l'échelle régionale et départementale, avant d'isoler le cas emblématique de l'Outre-mer et de le relier à des indicateurs sociaux fondamentaux que nous réutiliserons par la suite. Ces analyses spatiales, couplées à des métadonnées Insee, préparent le terrain pour la suite : une exploration plus fine aux niveaux de la circonscription

III - Analyse circonscriptionnelle et facteurs socio-économiques

III.1. Analyse en composantes principales (ACP) des indicateurs socio-économiques

Pour éclairer la diversité des profils socio-économiques des circonscriptions, nous avons conduit une ACP sur sept variables clés issues des statistiques Insee. Ces variables couvrent l'emploi, le chômage, le niveau de diplôme, la structure de l'emploi (cadres, employés, ouvriers) et l'âge moyen de la population. Ces variables n'ont pas été choisies au hasard car ce sont celles qui reviennent le plus souvent dans les actualités lors des périodes politiques comme l'élection présidentielle. Nous allons alors les étudier afin de voir si ces dernières expliquent les suffrages.



Nuage de points

Après centrage-réduction, l'ACP révèle deux dimensions majeures (78 % de la variance cumulée)

Premier axe (51,3 %) :

Variables fortement positives : taux de chômage, part de la population peu ou pas diplômée, proportion d'emplois ouvriers et employés.

Variables fortement négatives : part de cadres, taux d'emploi.

Les circonscriptions situées à droite de cet axe correspondent à des territoires marqués par la précarité (chômage élevé, faible niveau de diplôme, forte proportion d'emplois manuels), tandis que celles à gauche sont plutôt dynamiques et qualifiées.

Deuxième axe (27,0 %) :

Variable la plus contributive : âge moyen de la population.

Cet axe distingue des circonscriptions à population plus âgée (points en haut du graphique) de celles à population plus jeune (points en bas), indépendamment du profil d'emploi.

on remarque non seulement le gros « nuage » central (territoires à profil socio-économique proche de la moyenne nationale), mais aussi plusieurs points isolés aux quatre « coins » du graphique :

En haut à droite :

Ces circonscriptions combinent forte précarité (chômage et peu-diplômés élevés, part d'ouvriers/emplois manuels importante) et population plutôt âgée.

Exemples probables : Outre-mer

En bas à droite

Profil également marqué par la précarité, mais avec une jeunesse démographique (faible âge moyen).

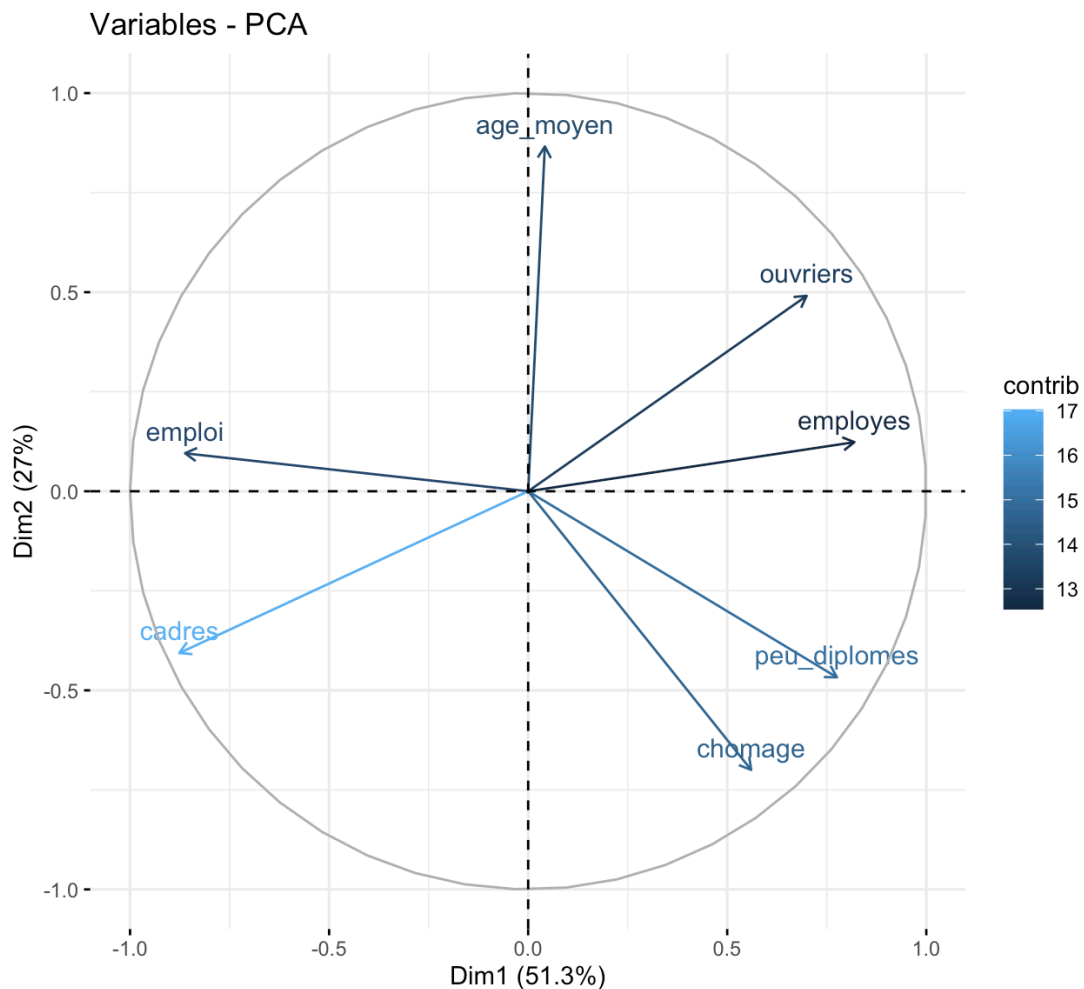
On y trouve peut-être de grandes banlieues ou zones périurbaines

En haut à gauche

Circonscriptions aisées (faible chômage, forte part de cadres/emplois qualifiés) et populations âgées : petits territoires ruraux ou villes historiques où la population est plus stable et qualifiée.

En bas à gauche

Territoires dynamiques et qualifiés mais jeunes, typiquement des métropoles universitaires ou des quartiers récents attirant de jeunes actifs très formés.



Cercle des corrélations socio-économique

*Le cercle des corrélations met en relation chaque indicateur avec les deux premiers axes :
L'angle et la distance depuis l'origine indiquent la force et le sens de la corrélation avec les axes.*

On y voit clairement des oppositions logiques "chômage/peu diplômés/ouvriers" vs "cadres/emploi", ainsi que la contribution quasi orthogonale de l'« âge moyen ».

Dim1 (51,3 %) : gradient qualification <-> précarité

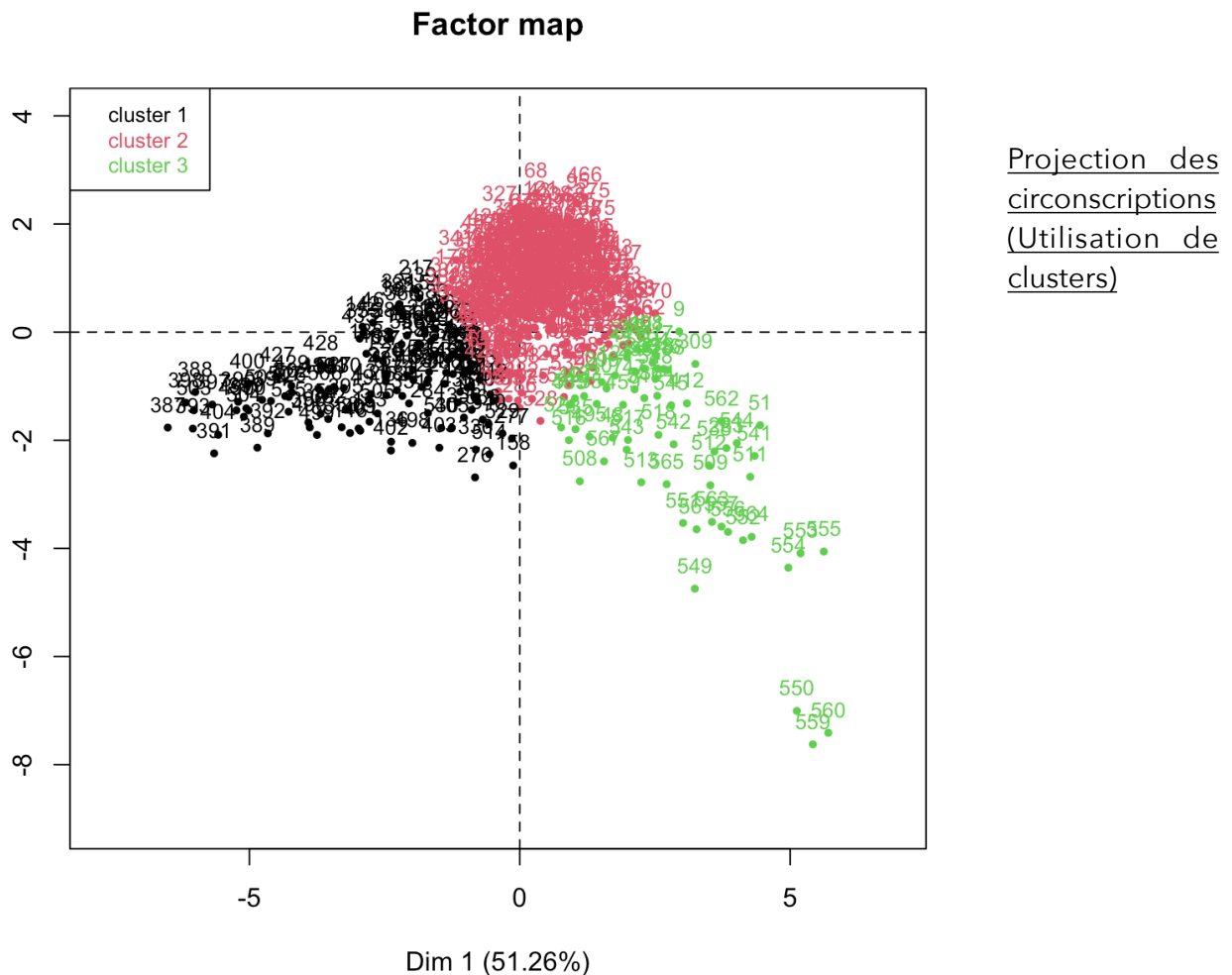
À droite, les vecteurs chômage, peu_diplômés et ouvriers pointent ensemble, indiquant qu'ils co-varient fortement : plus une circonscription se caractérise par un taux de chômage élevé et une forte proportion de travailleurs non qualifiés, plus elle se positionne sur le côté positif de Dim1.

À l'opposé (côté négatif), le vecteur cadres signale les zones où la part de cadres et de professions « nécessitant des diplômes » est élevée : ces circonscriptions se distinguent par un profil socio-professionnel plus favorisé.

Dim2 (27,0 %) : gradient vieillissement <-> jeunesse

Le vecteur « âge_moyen » pointe vers le haut : les circonscriptions à population plus âgée se retrouvent en haut du graphique.

Les autres indicateurs (emploi, employés) présentent des contributions plus modestes et se situent près de l'origine ou légèrement en dessous, suggérant un lien secondaire avec la dimension « âge ».



En projetant chaque circonscription sur Dim1 × Dim2, puis en appliquant un clustering, nous obtenons trois groupes socio-économiques distincts :

Cluster 1: ●

Caractéristiques : circonscriptions où ni la précarité ni la qualification extrême ne dominent ; âge moyen proche de la moyenne nationale.

Cluster 2: ●

Caractéristiques : fortes valeurs de « chômage », « peu_diplômés » et « âge_moyen »

Cluster 3: ●

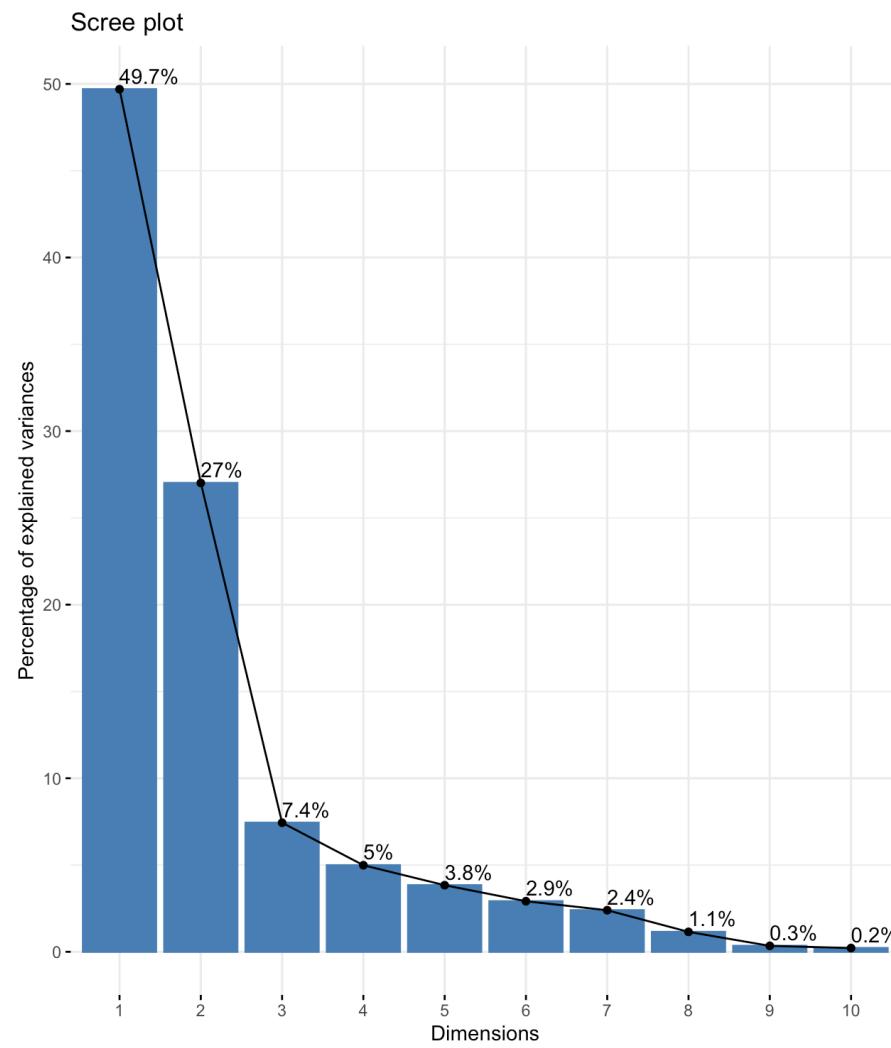
Caractéristiques : fortes valeurs de « chômage » et « peu_diplômés » mais faibles sur l'axe âge (population plus jeunes).

La segmentation en trois clusters reflète bien la dualité qualification/précarité, enrichie d'une dimension d'âge.

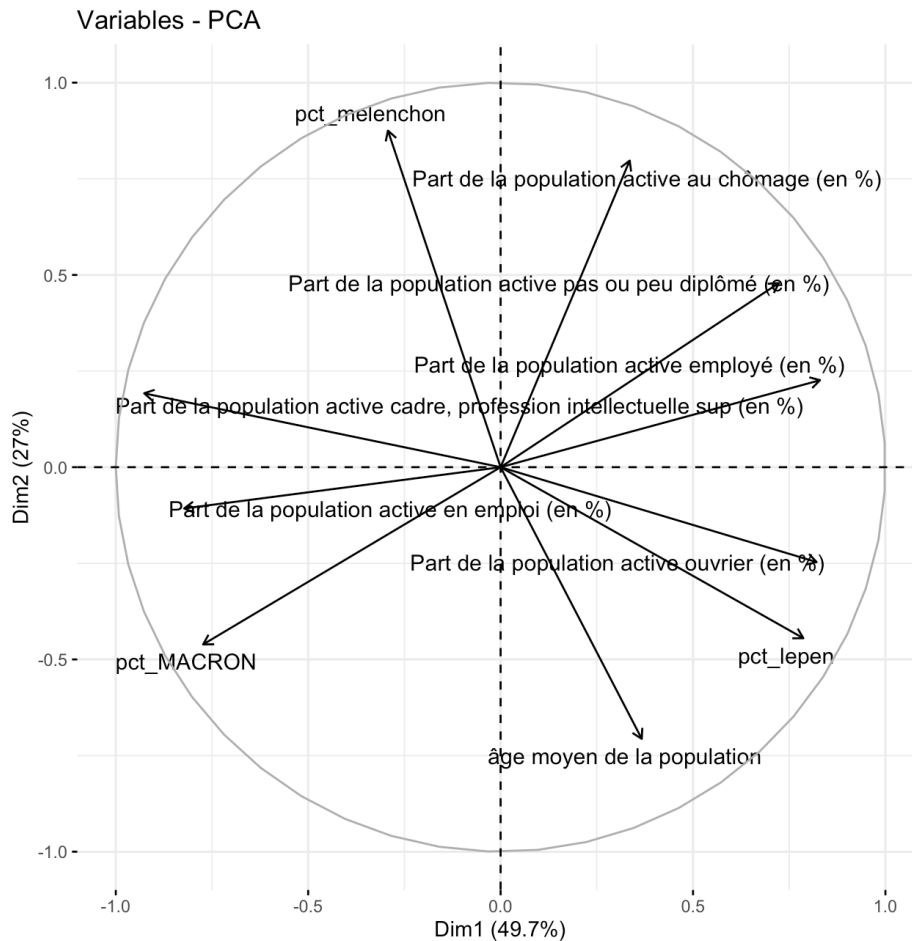
Ce typage permet de formuler des hypothèses testables : par exemple, on s'attend à ce que les circonscriptions précaires (Clusters 3 et 2) présentent des scoresLePen et scoresMelenchon plus élevés, tandis que les circonscriptions qualifiées (Cluster 1 et zones négatives de Dim1) favorisent davantage Macron.

III.2 ACP conjointe des indicateurs socio-économiques et des scores électoraux

Jusqu'ici, nous avons isolé la structure interne des variables socio-économiques. Pour comprendre comment ces profils s'associent aux intentions de vote, nous incluons maintenant, dans la même ACP, les scores (%) de Macron, Le Pen et Mélenchon, de façon à visualiser simultanément la position de chaque circonscription selon son contexte socio-économique et son vote.



La figure ci-dessus montre que les deux premiers axes expliquent à eux seuls 76,7 % de la variance (Dim 1 : 49,7 %, Dim 2 : 27,0 %) un peu comme sur notre première ACP

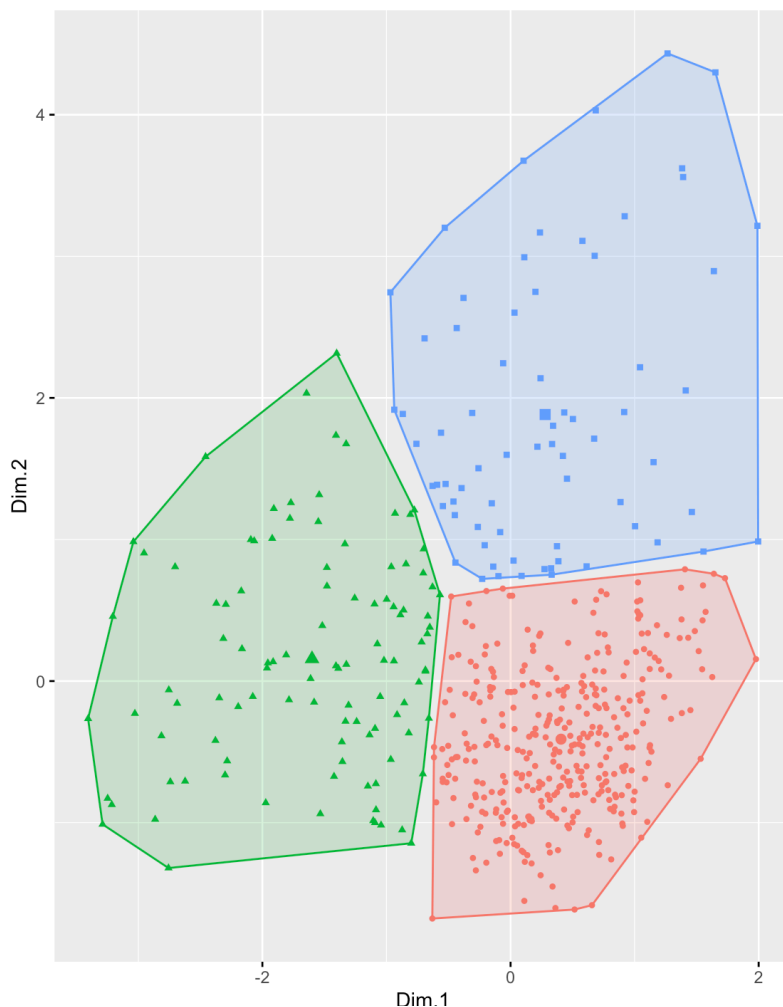


Cercle des corrélations socio-économiques et des scores électoraux

Le cercle des corrélations garde ici les mêmes vecteurs pour les indicateurs, cependant nous avons également ajouté des scores électoraux des candidats que nous allons interpréter.

- Pct_MACRON se projette nettement du côté négatif de Dim 1 (alignement avec les « cadres » et « l'emploi »), signe que Macron obtient ses meilleurs scores dans les circonscriptions « qualifiées » où avec un « haut niveau de vie ».
- Pct_lepen pointe vers le quadrant inférieur droit: Le Pen performe particulièrement dans les territoires à forte précarité et à population.
- Pct_melenchon se situe dans la partie supérieure gauche (négatif Dim 1, positif Dim 2) : Mélenchon réussit davantage là où la précarité est modérée à élevée, mais dans des territoires plus âgés et moins qualifiés que ceux de Macron.

Clusters de circonscriptions selon ACP + k-means



NB: (J'ai eu un problème sur les couleurs et les numéros des clusters)
la rectification par rapport à la partie précédente, dans nos suppositions:

Macron est cluster 2 (ancien

●)

Mélenchon est cluster 3 (ancien

●)

Le Pen est cluster 1 (ancien ●)

Ce clustering nous renvoie cette figure et nous obtenons les résultats ci-dessous:

Vers la gauche : circonscriptions « **Macron-friendly** »

Vers le bas-droite : circonscriptions « **Le Pen-friendly** »

Vers le haut-gauche : circonscriptions « **Mélenchon-friendly** »

L'intégration des scores dans l'ACP confirme que le vote suit en partie la géographie socio-économique. Chaque candidat s'ancre dans un profil de circonscription cohérent :

- Macron dans les zones diplômées et dynamiques,
- Le Pen dans un terreau de précarité modérée et de population jeune,
- Mélenchon dans des espaces marqués par la précarité et le vieillissement.

En effet, les clusters 1 et 3, identifiés comme précaires, correspondent aux profils de circonscriptions dans lesquelles les régressions révèlent une part significative du vote Le Pen et Mélenchon. Cela valide l'intuition qualitative par des estimations quantitatives robustes.

Globalement, l'affirmation se confirme pour Macron et Le Pen, mais elle mérite d'être nuancée pour Mélenchon car il a un score très homogène, et ne se repose pas que sur la population vieillissante en réalité.. (Chatgpt).

III.3 Régressions linéaires : vers une modélisation des parts de voix

MODÈLE « SOCIO-ÉCO SEUL

Nous ajustons pour chaque candidat le modèle sur l'ensemble des circonscriptions.

Call:

```
lm(formula = pct_MACRON ~ age_moyen + emploi + chomage + peu_diplomes +  
  cadres + employes + ouvriers, data = df_num)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.7329	-1.4793	-0.0658	1.3258	7.2394

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	31.80110	5.36923	5.923	5.71e-09 ***
age_moyen	-0.13646	0.04806	-2.840	0.00469 **
emploi	-0.21584	0.05095	-4.236	2.68e-05 ***
chomage	-1.67606	0.13322	-12.581	< 2e-16 ***
peu_diplomes	-0.43690	0.04114	-10.620	< 2e-16 ***
cadres	0.48628	0.04643	10.474	< 2e-16 ***
employes	0.12789	0.07378	1.733	0.08363 .
ouvriers	0.38590	0.04543	8.494	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.235 on 527 degrees of freedom

Multiple R-squared: 0.7567, Adjusted R-squared: 0.7535

F-statistic: 234.2 on 7 and 527 DF, p-value: < 2.2e-16

Régression Macron

Call:

```
lm(formula = pct_lepen ~ age_moyen + emploi + chomage + peu_diplomes +  
  cadres + employes + ouvriers, data = df_num)
```

Residuals:

Min	1Q	Median	3Q	Max
-10.6105	-2.5237	-0.2446	2.7239	12.3653

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.68850	9.04411	-0.297	0.766381
age_moyen	0.43813	0.08095	5.412	9.46e-08 ***
emploi	0.09192	0.08582	1.071	0.284659
chomage	0.23115	0.22440	1.030	0.303441
peu_diplomes	-0.01042	0.06930	-0.150	0.880496
cadres	-0.32724	0.07820	-4.185	3.35e-05 ***
employes	-0.12173	0.12428	-0.979	0.327789
ouvriers	0.25452	0.07653	3.326	0.000943 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.764 on 527 degrees of freedom

Multiple R-squared: 0.6419, Adjusted R-squared: 0.6372

F-statistic: 135 on 7 and 527 DF, p-value: < 2.2e-16

Régression Le Pen

```
Call:
lm(formula = pct_melenchon ~ age_moyen + emploi + chomage + peu_diplomes +
    cadres + employes + ouvriers, data = df_num)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-8.970 -2.576 -0.009  2.283 12.512
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  38.31124    8.86797   4.320 1.86e-05 ***
age_moyen    -0.61590    0.07937  -7.759 4.45e-14 ***
emploi        0.18407    0.08415   2.187  0.0292 *
chomage       1.48071    0.22003   6.730 4.45e-11 ***
peu_diplomes  0.33931    0.06795   4.994 8.06e-07 ***
cadres       -0.19302    0.07668  -2.517  0.0121 *
employes     -0.11590    0.12186  -0.951  0.3420
ouvriers     -0.52721    0.07504  -7.026 6.61e-12 ***
```

Régression Mélenchon

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.691 on 527 degrees of freedom
Multiple R-squared:  0.5971,    Adjusted R-squared:  0.5917
F-statistic: 111.6 on 7 and 527 DF,  p-value: < 2.2e-16
```

Ces premiers modèles expliquent 59 % à 76 % de la variance des parts de voix c'est pas mal mais on peut sûrement faire mieux, nous allons donc enrichir les équations précédentes de quatre variables additionnelles afin d'augmenter la valeurs de R carré:

log_vacants : part des logements vacants
log_old : part des logements construits<1990
taux_pauvre : taux de pauvreté (seuil 60 %),
revenu_med : niveau de vie médian.

on obtient donc:

```
Residuals:
    Min       1Q   Median       3Q      Max
-6.4248 -1.3651 -0.0992  1.3731  6.5806
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.807e+01  5.491e+00   5.112 4.49e-07 ***
age_moyen    -5.471e-02  5.352e-02  -1.022  0.30716
emploi       -2.737e-01  5.205e-02  -5.259 2.12e-07 ***
chomage      -1.246e+00  1.614e-01  -7.720 5.97e-14 ***
peu_diplo    -3.596e-01  4.887e-02  -7.357 7.30e-13 ***
cadres       4.928e-01  5.987e-02   8.230 1.50e-15 ***
employes     1.088e-01  7.433e-02   1.464  0.14378
ouvriers     4.141e-01  4.899e-02   8.453 2.83e-16 ***
log_vacants  -1.741e-01  5.870e-02  -2.966  0.00316 **
log_old      -2.823e-02  1.726e-02  -1.635  0.10263
taux_pauvre  -8.890e-02  5.493e-02  -1.619  0.10616
revenu_med    1.581e-04  8.591e-05   1.840  0.06634 .
```

Régression Macron enrichie

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.152 on 523 degrees of freedom
Multiple R-squared:  0.7762,    Adjusted R-squared:  0.7715
F-statistic: 164.9 on 11 and 523 DF,  p-value: < 2.2e-16
```

Call:

```
lm(formula = pct_lepen ~ age_moyen + emploi + chomage + peu_diplo +  
    cadres + employes + ouvriers + log_vacants + log_old + taux_pauvre +  
    revenu_med, data = df_reg)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-8.3672	-1.9179	-0.1755	2.0713	13.8655

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.6350467	8.3036168	1.281	0.200841
age_moyen	0.1574772	0.0809409	1.946	0.052240 .
emploi	-0.2161600	0.0787111	-2.746	0.006236 **
chomage	2.1287142	0.2440197	8.724	< 2e-16 ***
peu_diplo	0.2682255	0.0739079	3.629	0.000312 ***
cadres	-0.5815797	0.0905450	-6.423	3.00e-10 ***
employes	-0.1931482	0.1124021	-1.718	0.086321 .
ouvriers	-0.0107511	0.0740866	-0.145	0.884676
log_vacants	0.2570131	0.0887666	2.895	0.003945 **
log_old	0.0527026	0.0261054	2.019	0.044014 *
taux_pauvre	-0.7246798	0.0830681	-8.724	< 2e-16 ***
revenu_med	0.0005308	0.0001299	4.086	5.09e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.254 on 523 degrees of freedom

Multiple R-squared: 0.7344, Adjusted R-squared: 0.7289

F-statistic: 131.5 on 11 and 523 DF, p-value: < 2.2e-16

Régression Le Pen enrichie

Call:

```
lm(formula = pct_melenchon ~ age_moyen + emploi + chomage + peu_diplo +  
    cadres + employes + ouvriers + log_vacants + log_old + taux_pauvre +  
    revenu_med, data = df_reg)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-6.6872	-1.7938	-0.0315	1.6614	9.7883

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	34.1689917	7.0137044	4.872	1.47e-06 ***
age_moyen	-0.2822276	0.0683673	-4.128	4.26e-05 ***
emploi	0.5185812	0.0664839	7.800	3.38e-14 ***
chomage	-0.7505898	0.2061128	-3.642	0.000298 ***
peu_diplo	0.1421302	0.0624268	2.277	0.023204 *
cadres	0.3313780	0.0764794	4.333	1.77e-05 ***
employes	0.0429963	0.0949412	0.453	0.650828
ouvriers	-0.1984457	0.0625777	-3.171	0.001607 **
log_vacants	-0.1609541	0.0749773	-2.147	0.032276 *
log_old	-0.0716095	0.0220501	-3.248	0.001239 **
taux_pauvre	0.5632839	0.0701640	8.028	6.58e-15 ***
revenu_med	-0.0012920	0.0001097	-11.774	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.748 on 523 degrees of freedom

Multiple R-squared: 0.7783, Adjusted R-squared: 0.7736

F-statistic: 166.9 on 11 and 523 DF, p-value: < 2.2e-16

Régression Mélenchon enrichie

On observe une amélioration de l'ajustement :

Le R^2 ajusté progresse de manière notable pour Le Pen (+0,09), et plus modestement pour Macron (+0,02) et Mélenchon (+0,03).

Le vote Le Pen apparaît particulièrement sensible aux conditions de logement et de pauvreté, tandis que Mélenchon est corrélé à la pauvreté et aux faibles revenus médians. Un taux élevé de logements vacants pèse négativement sur le score Macron et sur Mélenchon, mais renforce Le Pen.

Logements anciens : davantage de parc antérieur à 1990 se traduit par un vote plus marqué pour Le Pen et un relatif désintérêt pour Macron/Mélenchon.

L'ajout des indicateurs de logement et de pauvreté affine sensiblement notre cartographie des déterminants du vote : on comprend mieux la fracture résidentielle.

Ces régressions permettent également de confirmer les résultats obtenus nos ACP réalisée au préalable, dans notre seconde ACP, nous avons intégré Pct_MACRON, Pct_LE PEN et Pct_MÉLENCHON parmi les variables (les scores des candidats) :

Les vecteurs de vote se positionnaient dans le cercle de corrélations exactement là où les indicateurs socio-éco définissaient Dim1 et Dim2.

Par exemple, le vecteur « Pct_MACRON » était très opposé à « chômage » et « peu_diplômés » (Dim1 négatif), tandis que « Pct_LE PEN » était plutôt aligné sur « âge_moyen » (Dim2 positif).

Les régressions linéaires viennent quantifier ces mêmes relations :

Les corrélations fortes observées en ACP entre « chômage »/« peu diplômés » et « Pct_Mélenchon » se traduisent par des $\beta_3, \beta_4 > 0$ et très significatifs.

L'opposition nette en ACP entre « cadres » et « Pct_Le Pen » apparaît comme $\beta_5 < 0$ très significatif. L'effet d'« âge moyen » sur « Pct_Le Pen » (Dim2) est chiffré par un $\beta_1 > 0$ très significatif.

En somme, nos régressions sont en accord avec les enseignements de l'ACP.

III.4 Régressions linéaires « tranches d'âge » (bonus)

Pour clore notre exploration, nous proposons un « bonus » : au-delà de l'âge moyen, essayons de voir comment la forme de la distribution d'âge d'une circonscription explique-t-elle le vote ? L'introduction des variables d'âge (moyenne et quantiles) vise à saisir deux phénomènes complémentaires :

- un effet d'âge « structurel », lié à la distribution démographique et un effet « générationnel », reflétant des attitudes politiques spécifiques à certaines tranches (jeunes adultes, seniors).

Nous modélisons donc, pour chaque candidat, la part de voix (pct_candidat) en fonction de : âge moyen

age_q10, age_q25, age_med, age_q75, age_q90 : respectivement 10^e, 25^e, 50^e, 75^e et 90^e centiles

R² très hétérogènes :

Macron (R²≈0.19) est faiblement expliqué par l'âge seul ;

Le Pen & Mélenchon (R²≈0.56) voient leur score bien prédit par la structure d'âge.

```
Macron - R² : 0.192 | R² ajusté : 0.183
LePen - R² : 0.557 | R² ajusté : 0.552
Melenchon - R² : 0.559 | R² ajusté : 0.554
```

```
> # 6) Résultats synthétiques
> tidy(lm_mac_ageq, conf.int = TRUE) # effet des tranches d'âge sur Macron
# A tibble: 7 x 7
  term          estimate std.error statistic  p.value conf.low conf.high
  <chr>         <dbl>     <dbl>     <dbl>    <dbl>   <dbl>   <dbl>
1 (Intercept) -10.3      7.79      -1.32 1.87e- 1 -25.6    5.01
2 age_mean      4.92     1.23       4.01 6.94e- 5  2.51    7.34
3 age_q10       1.34     0.336     3.99 7.55e- 5  0.681    2.00
4 age_q25      -1.40     0.256     -5.48 6.70e- 8 -1.91   -0.900
5 age_med      -0.934    0.389     -2.40 1.66e- 2 -1.70   -0.170
6 age_q75      -2.57     0.344     -7.46 3.56e-13 -3.24   -1.89
7 age_q90       0.522    0.288      1.81 7.05e- 2 -0.0438  1.09
> tidy(lm_lep_ageq, conf.int = TRUE) # idem pour Le Pen
# A tibble: 7 x 7
  term          estimate std.error statistic  p.value conf.low conf.high
  <chr>         <dbl>     <dbl>     <dbl>    <dbl>   <dbl>   <dbl>
1 (Intercept)   4.81     8.02       0.600 5.49e- 1 -10.9    20.6
2 age_mean       2.44     1.26       1.93 5.39e- 2 -0.0413  4.92
3 age_q10      -0.711    0.346     -2.06 4.04e- 2 -1.39   -0.0313
4 age_q25      -1.52     0.263     -5.78 1.30e- 8 -2.04   -1.00
5 age_med      -0.227    0.400     -0.568 5.70e- 1 -1.01    0.559
6 age_q75       1.67     0.354       4.71 3.15e- 6  0.971    2.36
7 age_q90      -1.92     0.296     -6.49 2.00e-10 -2.51   -1.34
> tidy(lm_mel_ageq, conf.int = TRUE) # idem pour Mélenchon
# A tibble: 7 x 7
  term          estimate std.error statistic  p.value conf.low conf.high
  <chr>         <dbl>     <dbl>     <dbl>    <dbl>   <dbl>   <dbl>
1 (Intercept)  85.9      7.39      11.6 5.33e-28 71.4    100.
2 age_mean     -5.52     1.16      -4.75 2.65e- 6 -7.81   -3.24
3 age_q10     -0.454    0.319     -1.42 1.55e- 1 -1.08    0.172
4 age_q25      2.33     0.243      9.59 3.35e-20  1.85    2.81
5 age_med      0.852    0.369      2.31 2.12e- 2  0.128    1.58
6 age_q75      0.638    0.326      1.96 5.09e- 2 -0.00238 1.28
7 age_q90      0.550    0.273      2.01 4.46e- 2  0.0134    1.09
```

Que nous apportent ces régressions?

Macron : profil « extrêmes »

Macron capte particulièrement les électorats des deux extrêmes d'âge :

les très jeunes (forte part sous $Q_{10}+1,34$)

et, dans une moindre mesure, les très âgés ($Q_{90}+0,52$)

En revanche, il est affaibli dans les zones à population globalement adultes/moyennes ($Q_{25} \rightarrow Q_{75}$).

L'effet le plus « lourd » (en valeur absolue) reste la part de la population entre Q_{25} et Q_{75} : là où l'âge est « normal », il perd le plus de voix.

Le Pen : profil « classes d'âge supérieures »

Marine Le Pen tire son électorat principalement des classes d'âge « senior », mais pas des plus âgés (Q_{90}).

Ses meilleurs résultats se situent là où la proportion de 65-75 ans (autour de Q_{75}) est forte.

Elle est nettement pénalisée dans les zones où la part de très jeunes et de très vieux est importante.

Mélenchon : profil « jeunes adultes et classes intermédiaires »

Jean-Luc Mélenchon est très sensible à la présence de jeunes adultes (Q_{25}) et de populations actives (Q_{75}).

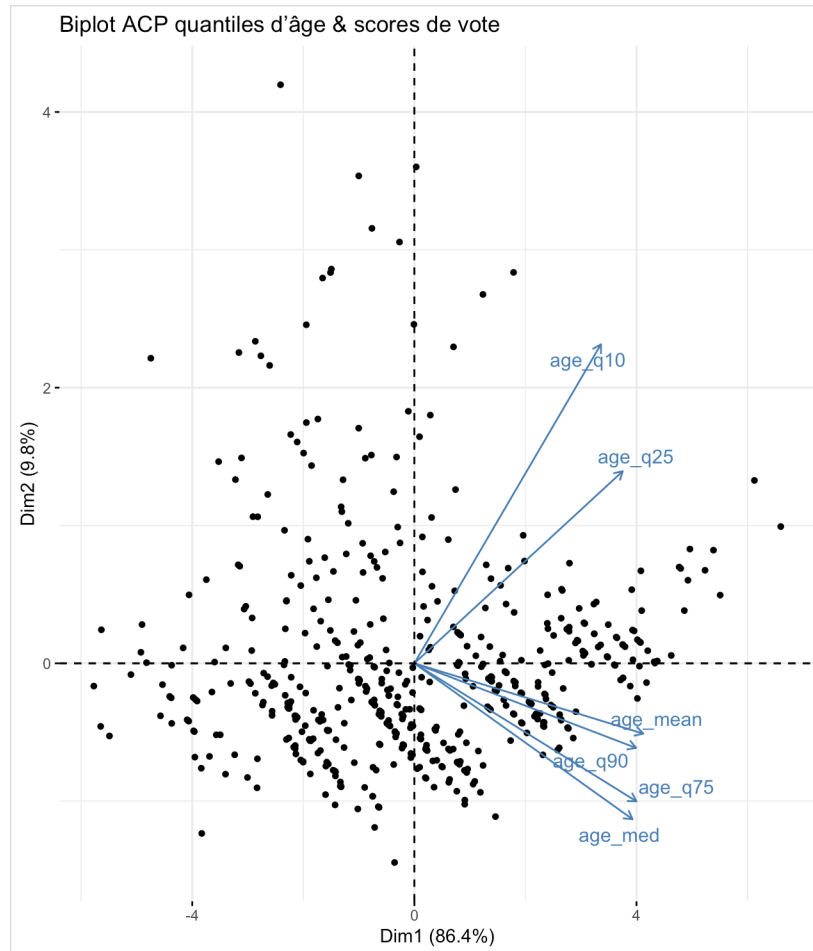
À l'inverse, dès que la moyenne d'âge grimpe, son score s'effondre (estimateur négatif très significatif).

L'effet du 10^e percentile (Q_{10}) n'est pas significatif : il n'attire pas spécialement les tout jeunes.

C'est très intéressant pour ce candidat car les analyses précédentes nous disaient qu'il prenait beaucoup de voix sur des tranches plutôt âgées, or cette régression nous permet de voir qu'il a une forte part des 25 % plus jeunes : +2,33 (jeunes adultes)

À l'inverse, l'âge moyen de la circonscription a un effet négatif très fort (-5,52 pts), ce qui signifie qu'au-delà de ces segments de population active, Mélenchon perd du terrain. Ce qui explique sûrement nos résultats obtenus précédemment.

On a également réaliser une ACP sur ce même thème



Le nuage de point des quantiles d'âge révèle en fait une structure très simple ce qui était plus ou moins attendu.

Un seul axe dominant (Dim 1 = 86 % de la variance) tous les quantiles (q10, q25, médiane, q75, q90) pointent dans le même sens : ils co-varient quasi parfaitement. Il n'y a qu'un gradient unique « jeunes <-> âgés » ; les circonscriptions se répartissent selon leur position globale sur cet axe. Un deuxième axe quasi anecdotique (Dim 2 = 9,8 %). Il oppose légèrement les quantiles inférieurs (q10, q25) aux quantiles supérieurs (q75, q90), mais son inertie est très faible. On y perçoit à peine la dissymétrie ou l'asymétrie de la distribution d'âge (quelques circonscriptions où la proportion de très jeunes diffère subtilement du reste), On a pas de groupes statistiques spontanés contrairement à l'ACP socio-économique, ici il n'y a pas de sous-ensembles (clusters) de circonscriptions : on n'identifie ni « quart de population particulièrement âgé » distinct, ni « quart particulièrement jeune » isolé.

Que peut-on en tirer ?

On peut en déduire que l'âge moyen résume quasiment toute l'information démographique d'âge : ajouter les quantiles n'apporte quasiment aucune dimension nouvelle. Donc sur ce point le choix de prendre uniquement l'âge moyen lors des études précédentes était judicieux. Limite méthodologique : si l'on veut vraiment creuser la dimension « âge », il faudrait introduire d'autres variables (dispersion...) ou travailler à un niveau plus fin (par exemple la proportion d'adhérents à certaines classes d'âge), mais pas se cantonner aux quantiles seuls.

CONCLUSION

Nous avons montré que la géographie électorale française s'inscrit avant tout dans un continuum « qualification vs précarité », modulé par un gradient d'âge, et que l'ajout d'indicateurs de logement et de pauvreté affine notre compréhension, en particulier pour les votes populistes.

D'un point de vue théorique, ces résultats confortent et prolongent les travaux de Lipset & Rokkan (1967) ainsi que les analyses récentes de Desage & Guérin-Pace (2005) sur le vieillissement démographique.

Facteurs structurants du vote

nos ACP mettent en lumière un axe principal opposant territoires à forte proportion de cadres et taux d'emploi élevé à territoires marqués par le chômage, la faible qualification et une part importante d'ouvriers. Les régressions confirment qu'une plus grande qualification (cadres, emploi) augmente significativement le score de Macron, tandis qu'une précarité socioprofessionnelle renforce Mélenchon et, dans une moindre mesure, Le Pen.

Gradient démographique : l'âge moyen se distingue en ACP comme second facteur, et les modèles d'âge montrent que les jeunes adultes (25 % inférieurs) soutiennent fortement Mélenchon, que les seniors (jusqu'au 75^e percentile) préfèrent Le Pen, et que Macron profite surtout des extrêmes d'âge (très jeunes et très vieux).

Variations selon l'échelle et l'espace ultramarin

à l'échelle nationale, trois candidats concentrent 73 % des suffrages exprimés, mais leur implantation diffère ; Macron domine en Île-de-France et Outre-mer éduqué, Le Pen dans le Nord-Est rural, Mélenchon dans le Sud-Ouest et les zones populaires.

Départements : l'analyse fine confirme ces pôles de force et révèle des écarts extrêmes (Mayotte : +25 pts Le Pen vs Macron ; Français hors de France : +25 pts Macron vs Mélenchon).

Outre-mer : les DOM cumulent forte abstention et précarité (33,7 % de pauvreté vs 14,6 % en Métropole), ce qui explique l'extraordinaire succès de Mélenchon dans les DOM les plus paupérisés et celui de Macron dans les collectivités ultramarines plus diplômées.

Apports des méthodes multivariées

ACP : offre un cadre synthétique pour dégager les grands gradients socio-économiques et démographiques.

Clustering : valide la segmentation en trois profils (« qualifiés », « précaires jeunes », « précaires âgés ») et rattache chaque candidat à un profil de circonscription.

Régressions linéaires : quantifient l'effet propre de chaque variable, démontrent l'importance du logement et de la pauvreté pour Le Pen et Mélenchon, et distinguent l'effet d'âge structurel de l'effet générationnel

Nos modèles actuels présentent toutefois les limites classiques de l'analyse agrégée : un biais écologique et une sensibilité aux spécifications linéaires.

Pour surmonter ces contraintes, nous recommandons : l'exploitation de données de bureau de vote plus fines pour réduire l'hétérogénéité intra-circonscriptionnelle ; In fine, ce travail apporte un éclairage nouveau pour moi sur les ancrages socio-économiques et démographiques du vote présidentiel.

Biblio et Webographie

[1] resultats-par-niveau-fe-t1-france-entiere

[2][https://fr.wikipedia.org/wiki/](https://fr.wikipedia.org/wiki/Abstention_électorale_en_France#Analyse_des_élections_de_2017)

Abstention électorale en France#Analyse des élections de 2017

[3] resultats-par-niveau-reg-t1-france-entiere.xlsx.

Lipset & Rokkan (1967) - *Structures de clivages, systèmes de partis et alignement des électeurs* Blais &

Nadeau (1996) - *Measuring strategic voting: A two-step procedure*

Desage & Guérin-Pace (2005) - *Sociologie des réformes et des institutions intercommunales* Jackman

& Miller (1995) - *Voter turnout in the industrial democracies during the 1980s*

Merci de votre attention